Warn Levels: Ordering Data for Custom Filtration

Dr. Lukas Mandrake, Dr. Gary Doran – Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109



What You Need To Know

What's a Warn Level?

A number for each delivered OCO-2 sounding 0 to 19

> Estimates sounding contamination (lower is better)

WL = 0-5 is best quality data

WL 6-15 can be useful but has rising issues

WL 16-19 is likely and increasingly useless

Why can't I use a Quality Flag?

You can, but they aren't made for you: Data provider guesses which data "good enough"

- Passes inflexible, small % of total data -

What if you need 30% instead of 20%?

What if you want even more filtration to 10%?

How do Warn Levels help?

Alternative to Quality Flag: Data Ordering

Just provide users a data ordering

No good/bad decisions, no cutoff, no lost data

User decides how far into the ordering to use







How to Use Warn Levels?

- Decide requirements beforehand: how much data volume / coverage or scatter / error is needed or tolerable?

Monitor above statistics. 3 Stop when data volume / coverage are

2 Begin admitting WL = 0, 1, 2, ... into project.

You're done!

acceptable, or when scatter / error become

You now have a custom filter for your analysis

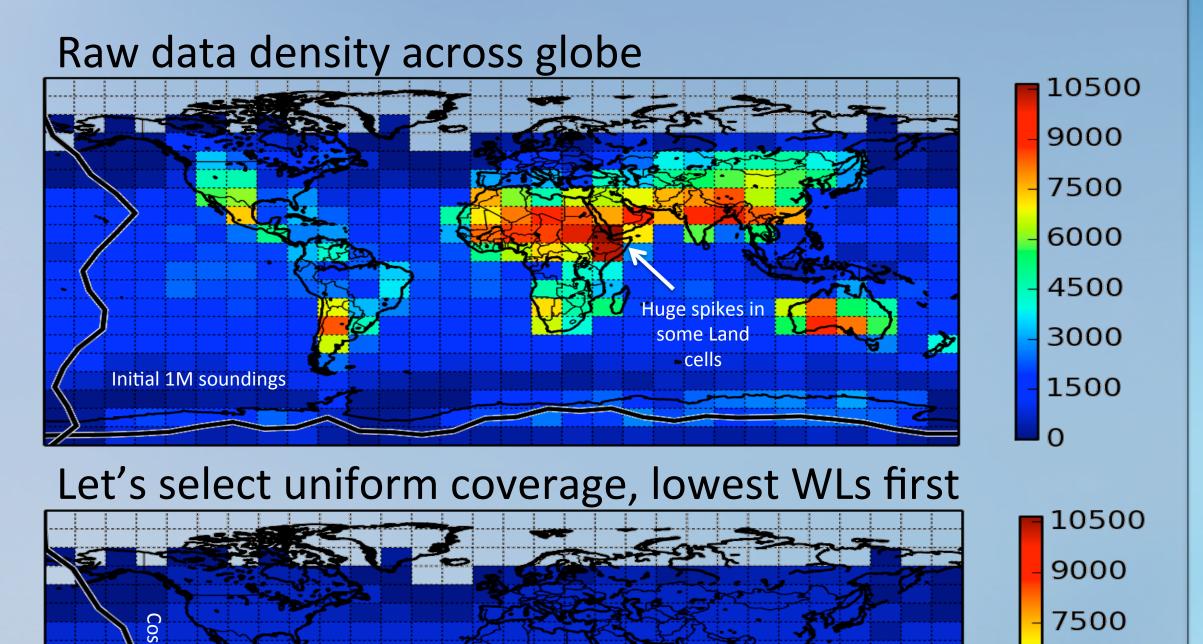
When collaborating:

intolerable.

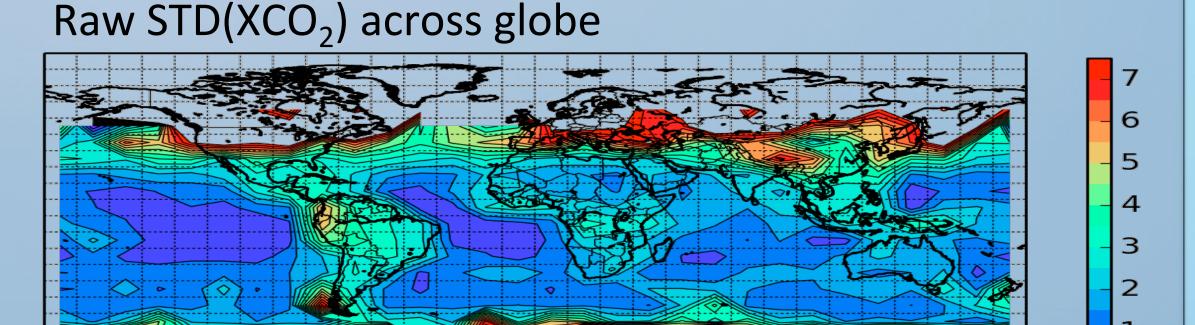
"I used Warn Levels < N" to define filtration strategy

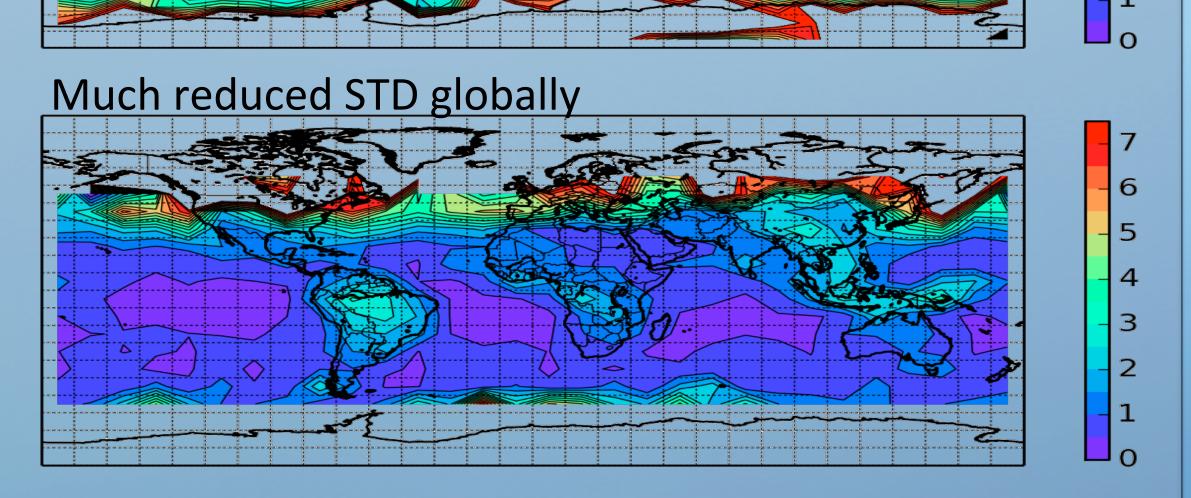
Watching Them Work

Selecting your dataset (oco-2 v5 data)

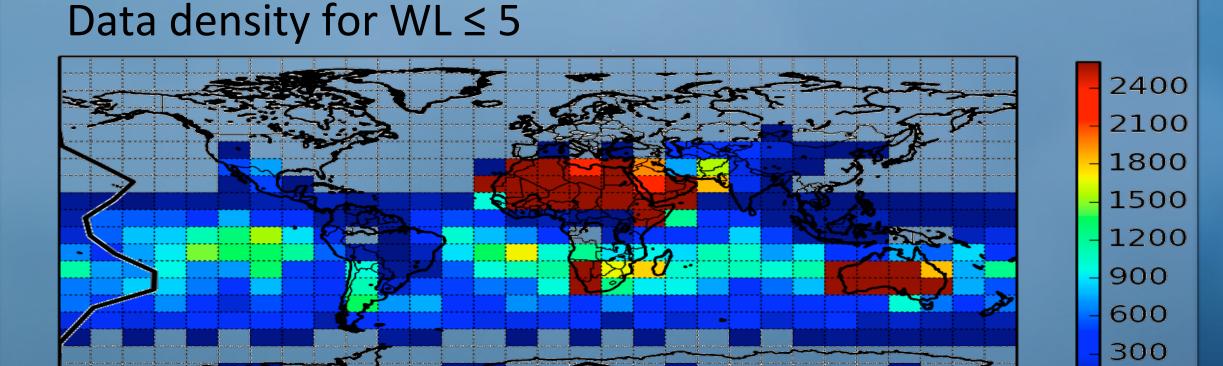


What did using lowest WL first buy you?





Want high-quality data, don't care where?



Warn Level Summary:

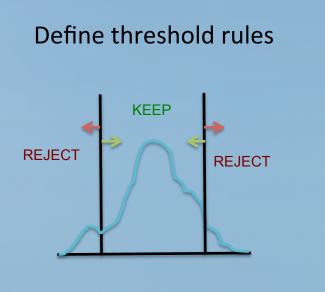
- Show where the highest quality data resides
- Intelligently order data for bin fill / user selection
- Reduce error metrics (such as STDEV)
- Easily communicate to colleagues which data used

How Are They Made?



DATA ORDERING GENETIC OPTIMIZATION Winner JPL Software of the Year

What do human experts do?

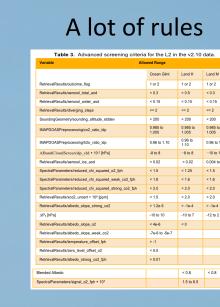


6000

4500

3000

1500





Much expert time & effort yields single filter based on many rules. Binary output: yes or no for each sounding.

DOGO -1- The Metrics

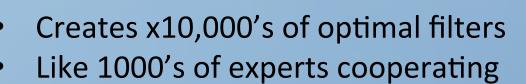
What DOGO minimizes when ordering the data Explored three:

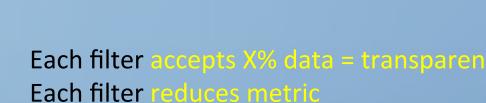
A DOGO Metric measures something unpleasant about the data

- Minimize STDEV(XCO2) in the southern hemisphere
- Minimize STDEV(XCO2) in individual Small Areas (see Wennberg)
- Minimize RMS difference with TCCON truth (where available)

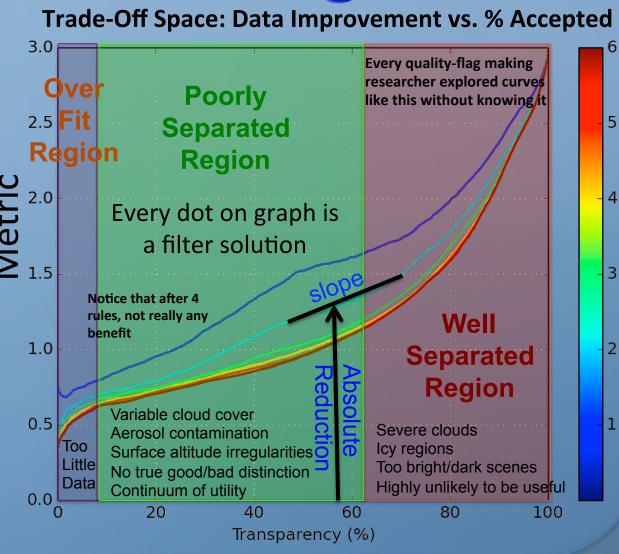
Final Warn Levels merge the knowledge from all 3 metrics

DOGO -2- The Genetic Algorithm



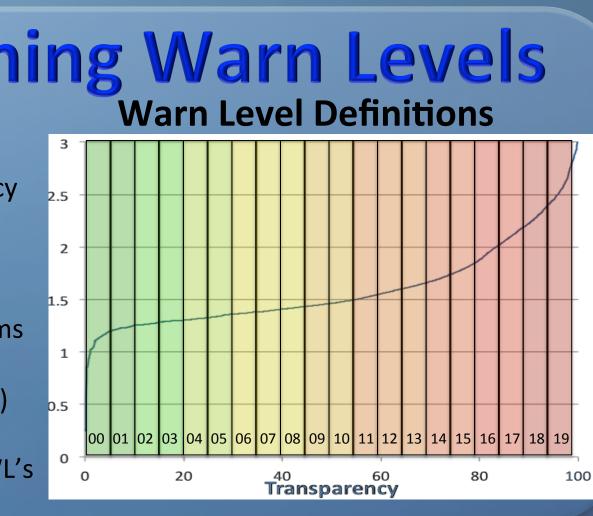


- Graph to produce Trade-off Curve $\sum_{1.5}$ Fundamental shape of data quality
- Each dot on graph is a valid filter



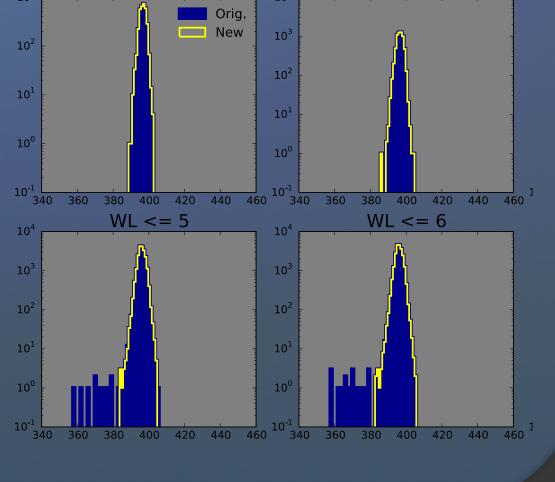
DOGO -3- Defining Warn Levels

- Use trade-off curves to create WL's
- WL's chop this graph every 5% transparency
- Must combine curves from all Metrics
- Discovers the features that predict problems
- Discovers how many there are (complexity)
- Discovers feature thresholds that define WL's



DOGO -4- Removing Outliers

- WL's are defined by improving "bulk statistics"
- Lone outliers that don't influence bulk metric may linger
- Identify outliers using model
- Re-categorize outliers to higher WL's
- Produces cleaner WL histograms
- Removes "surprise" outliers, especially at low WL's



WL <= 1

New in OCO-2 v7.1 data!